CIS 449 Assignment 5: I Want Candy!

Due: July 9 @ 11:59 PM

In this assignment, we are going to use the tools of the last two weeks to help us think about something a bit more fun: Martians manufacturing candy.



For those of you that don’t know, Mars (the company) makes the candy M&M’s. When thinking about distributional statistics, there are few better things that we could explore because there are inherently proportions, groupings, and numerical measurements that can be taken.

As such, we are going to use such a dataset that took three medium size bags of M&M’s, each of a different type:

* Plain
* Peanut
* Peanut Butter

M&M’s come in six different colors standardly:

* Blue
* Brown
* Green
* Orange
* Red
* Yellow

With the items in those bags, the color of each individual M&M was noted, and two measurements were taken:

* Diameter
* Weight

This data set was collected in the summer of 2008. Every M & M candy from three Medium Size bags of M & Ms was measured. One bag was of plain M & Ms, (14.0 oz. or 396.9 g), one bag was of peanut M & Ms (also 14.0 oz. or 396.9 g), and one bag was of peanut butter M & Ms (12.7 oz. or 360 g).

The variable diameter refers to the shortest distance from side to side at the candy's widest height when it is placed flat on the table with the "m" facing up. Put otherwise, when the candy is placed in that position, imagine taking horizontal cross-sections of the candy. They will be roughly elliptical. The diameter of the candy is the length of the minor axis of the largest such cross-sectional ellipse (which will generally be the cross-section at half the total height). As you might expect, this axis can be somewhat difficult to determine and was no doubt a source of measurement error, but this definition of diameter does correspond fairly well to the way that an M & M fits into a caliper.

Diameters were measured with a General Tools Ultratech Fraction+ Digital Fractional Caliper (claimed accurate up to plus or minus 0.02mm), and masses were measured with a MyWeigh Durascale 50 (claimed accurate up to plus or minus 0.01 g). The candies were measured in the order given in the data set, which although not entirely random was not intentionally systematic in any way (other than by type).

Check out this video to see how M&Ms are made: <https://www.youtube.com/watch?v=GtfkemNzG_I>

With this dataset, I want you to write an R script that will answer the following questions. When asked a yes or no question, find your answer using a statistical test.

1. What is the mean and standard deviation of the following cross sections of data for both diameter and mass: (20 points)
   1. Type
   2. Type+Color
2. What is the proportion of color per each type? (10 points)
3. Make a bar chart summarizing #2, colored appropriately. (5 points)
4. Is the color of the M&M’s evenly distributed in each type of bag? (5 points)
5. Are the M&M’s of different types the same weight? (5 points)
6. Are the M&M’s of different types the same diameter? (5 points)
7. Now we are going to make random bags of M&Ms of each type under the assumption that the color of the M&M has no bearing on its mass or diameter. Do this under two separate assumptions: (10 points)
   1. As a maximization of profit, do not exceed the weight of the bag posted, but do not leave room for another random candy
   2. Because we value the customer, make sure the bag at minimum has the weight posted, but do not kill our profits by putting any extra in (in other words, keep the M&M in the bag that makes it pass the weight limit and do not add any more)
8. Now we are going to revise the procedure: color matters. As you add M&Ms to the bag, you must first generate their color randomly and then use the parameters from that color (calculated in Step 1 B) to specify the values for weight and diameter randomly using a distribution. Use the same assumptions for your previous bag samples. (20 points)
9. Use your procedure in 8A to now try to create an excess bag – how many bags of M&M’s did you have to remove the extra M&M from to fill an additional bag? When you jettison the last M&M, put it in a bag you are filling up with the excess only. This gives you an idea of how much Mars would save (as a percentage) if they were to use strategy A as opposed to strategy B. Do this for each type of M&M. (20 points)

For #8 and #9 in particular, a hint is to try to create a picture of what is happening. That picture and description will be worth its weight in gold to you.

I want you to treat this as if you are trying to run the numbers for the head of Mars. As such, #9 is the thing that matters and it is built off the back of #1, #2, and #8 more so than anything else. The rest of the questions are worth smaller amounts than they otherwise might be to reflect that importance. Question 3 is just a visualization of #2; Questions #4-#6 are the questions that motivate the difference between #7 and #8. Logically for the business, #8 makes the most sense because the M&Ms are coated by different machines, and thus a color difference might exist. Each of those machines could have its own issue, so independent of what #4-#6 might suggest, still assume #8 is much more appropriate to use, though #7 is conceptually simpler.